Guidelines for implementing SAML authentication proxies for social media identity providers

Published Date:	13-06-2017
Revision:	1.0
Work Package:	JRA1
Document Code:	AARC-JRA1.4G
Document URL:	https://aarc-project.eu/wp-content/uploads/2017/03/AARC-JRA1.4G.pdf

Table of Contents

1 Introduction

One of the major goals of the blueprint architecture is to support users in research collaborations who do not have a federated identity via their home organisation. Moreover, there are cases in which an individual researcher is not affiliated with any of the traditional home organisations. To cater for these cases, the AARC blueprint architecture enables research communities and infrastructure providers to connect to identity providers that are not part of any of the eduGAIN participating federations. Such guest identity providers include social networks, which typically use OpenID Connect/OAuth2 for authentication and authorisation. This subsection provides recommendations and best practices for implementing authentication proxies that can connect social media identity providers with federated SAML 2.0 service providers.

2 General guidelines

The guidelines presented in this subsection have been defined based on experiences from multiple parties in the AARC project and have subsequently been discussed and tested through the SA1 pilot on using social media as guest identities for federated access [AARC-SA1-SCP].

- In order for the proxy to support the REFEDS Research and Scholarship [REFEDS-RS] attribute bundle, the RI/EI needs to make sure that the social authentication application (typically an OAuth2/OIDC client) is properly configured to request the required data elements from the social IdP. The RI/EI should therefore set up appropriate permissions and request scopes to allow users to authorise their social IdP to release information such as the shared user identifier and email address.
- In SAML, the recommended user identifier is the eduPersonUniqueId (ePUID) [I2-EPUI], which is a long-lived, non-reassignable, shared identifier. While ePUID is formatted like an email address, it is not intended to be a person's published email address or to be used as an email address. In fact, the released email address should never be used for the user's ePUID as social identities can have multiple email addresses at different points in time.
- In the case of OIDC-compliant social IdPs, the subject (sub [OIDC-Sub]) and issuer (iss [OIDC-Iss]) claims can be used together as a stable global identifier for the end user, since the sub claim is locally unique and never reassigned within the issuer for a particular end user. Therefore, the combination of the iss claim and the sub claim is appropriate for calculating a SAML ePUID. Any algorithm with the following properties can be used to calculate ePUIDs:

    - Distinct combinations of the iss and the sub claim MUST result in distinct ePUID values.
    - The algorithm MUST be deterministic.
    - The "uniqueId" portion MUST contain only alphanumeric characters (a-z, A-Z, 0-9).
    - The "uniqueId" portion MUST be less than or equal to 64 characters.

- The “scope” portion MUST be the administrative domain of the social IdP proxy where the identifier was created and assigned.
- The “scope” portion MAY contain any Unicode character.
- The “scope” portion MUST be less than or equal to 256 characters.

Based on the properties above, this guideline propose the following algorithm:

ePUID = SHA-256 (sub || iss || salt) || ‘@’ || scope

where the sub claim is concatenated with the iss claim and a static salt value. The concatenated string is then hashed using SHA-256. The result is then scoped at the administrative domain of the authentication proxy where the identifier was created and assigned.

- In the event that a non-shared (targeted) user identifier is released by the social IdP, then different SAML authentication proxies will receive distinct user identifier values for the same end user. This will result in distinct ePUID values, even if the same generation algorithm is being used. However, it should be noted that the services behind a given authentication proxy will still be able to identify users consistently, since the proxy-specific user identifier will remain the same.

3 Example implementations

This section describes how some of the most commonly used social identity profiles can be mapped to SAML attribute assertions. It should be noted that only user information which is relevant to the REFEDS Research and Scholarship (R&S) attribute bundle is covered here.

3.1 Google / OpenID Connect

Google's OAuth 2.0 APIs[1] can be used for both authentication and authorisation. This OAuth 2.0 implementation conforms to the OIDC specification and is OpenID Certified. Thus, information about the user can be retrieved from the UserInfo endpoint in OpenID Connect format by including the openid scope. The Claims[2] returned in the UserInfo Response can be mapped to SAML attributes as shown in Table 1. Note that

[1] https://developers.google.com/identity/protocols/OAuth2

[2] https://openid.net/specs/openid-connect-basic-1_0.html#StandardClaims

the included sub is the user identifier, which is unique among all Google accounts, persistent and non-reassignable[3]. The above apply to any OIDC-compliant IdP.

3.2 Facebook

Facebook allows retrieving user information through the /{user-id} Graph API endpoint[4], following an OAuth 2.0 flow for authentication and authorisation. The returned fields of the Facebook user profile can be mapped to SAML attributes as shown in Table 1. Note that the Graph API returns by default only the id field, i.e. a user ID unique to each app that cannot be used across different apps. Effectively, this is a targeted ID that can be mapped to a SAML 2.0 Persistent NameID/ePTID. In order to retrieve an anonymous, but unique identifier for the user that can be shared with third parties, the third_party_id field must be requested explicitly.

3.3 LinkedIn

LinkedIn relies on the OAuth 2.0 protocol for enabling authenticated access to its REST APIs that provide access to member data. More specifically, following a three-legged OAuth2 flow, LinkedIn user profile[5] information can be accessed through the /people/~ REST API endpoint. Table 1 shows how the returned user fields can be mapped to SAML attributes. Note that the id user field returned by LinkedIn is a unique identifying value for the user, which is linked to the specific application (targeted identifier).

3.4 Summary

Table 1 summarises the mappings to SAML attributes for some of the most commonly used social media user profiles.

SAML	Google / OIDC	Fecebook	LinkedIn
ePUID	sub || issuer	third_party_id (3) || 'facebook.com'	id || 'linkedin.com'

[3] https://developers.google.com/identity/protocols/OpenIDConnect#obtainuserinfo

[4] https://developers.facebook.com/docs/graph-api/reference/user

[5] https://developer.linkedin.com/docs/fields/basic-profile

displayName	name (1)	name	formatted-name (4)
givenName	given_name	first_name	first-name (4)
sn	family_name	last_name	last-name (4)
mail	email (2)	email	email-address (5)

Table 1 Mapping social identities to the REFEDS R&S attribute bundle.
(1) Provided when the request scope included the string "profile" or the ID token is returned from a token refresh
(2) Provided when the request scope included the string "email"
(3) Provided when the fields query parameter included the string “third_party_id”; otherwise only id is made available
(4) Provided when the LinkedIn application has requested the r_basicprofile member permission
(5) Provided when the LinkedIn application has requested the r_emailaddress member permission

4 References

[AARC-SA1-SCP] AARC wiki page: SocialIDCockpitPanel
https://wiki.geant.org/display/AARC/SocialIDCockpitPanel

[REFEDS-RS] REFEDS web page: Research and Scholarship Entity Category
https://refeds.org/category/research-and-scholarship

5 Glossary

API Application Program Interface
HTTP Hypertext Transfer Protocol
IdP Identity Provider
OIDC OpenID Connect
REST REpresentational State Transfer
SAML Security Assertion Markup Language
SP Service Provider